UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/644,815	08/21/2003	Jerome R. Bellegarda	P2989-908	6190

21839   7590   06/21/2007

BUCHANAN, INGERSOLL & ROONEY PC
POST OFFICE BOX 1404
ALEXANDRIA, VA 22313-1404

EXAMINER
DWIVEDI, MAHESH H

ART UNIT	PAPER NUMBER
2168	

MAIL DATE	DELIVERY MODE
06/21/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE <u>3</u> MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on <u>21 March 2007</u>.

2a)☒ This action is **FINAL**.   2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-11,13-28 and 30-47* is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-11, 13-28, and 30-47* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *21 August 2003* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All   b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1)☐ Notice of References Cited (PTO-892)

2)☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3)☐ Information Disclosure Statement(s) (PTO/SB/08)
    Paper No(s)/Mail Date _____.

4)☐ Interview Summary (PTO-413)
    Paper No(s)/Mail Date. _____.

5)☐ Notice of Informal Patent Application

6)☐ Other: _____.

DETAILED ACTION

Response to Amendment

1.   Receipt of Applicant's Amendment, filed on 03/21/2007, is acknowledged. The amendment includes the cancellation of claims 12 and 29, the addition of claims 37-47, and the amending of claims 9, 11, 12-15, 17, 21, 25, 27-28. 30, and 35.

Information Disclosure Statement

2.   The listing of references in the specification is not a proper information disclosure statement. 37 CFR 1.98(b) requires a list of all patents, publications, or other information submitted for consideration by the Office, and MPEP § 609.04(a) states, "the list may not be incorporated into the specification but must be submitted in a separate paper." Therefore, unless the references have been cited by the examiner on form PTO-892, they have not been considered.

Claim Objections

3.   The objections raised in the office action mailed on 12/21/2006 have been overcome by the applicant's amendments received on 03/21/2007.

Claim Rejections - 35 USC § 112

4.   The rejections raised in the office action mailed on 12/21/2006 have been overcome by the applicant's amendments received on 03/21/2007.

Claim Rejections - 35 USC § 103

5.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

This application currently names joint inventors. In considering patentability of the claims under 35 U.S.C. 103(a), the examiner presumes that the subject matter of the various claims was commonly owned at the time any inventions covered therein were made absent any evidence to the contrary. Applicant is advised of the obligation under 37 CFR 1.56 to point out the inventor and invention dates of each claim that was

not commonly owned at the time a later invention was made in order for the examiner to
consider the applicability of 35 U.S.C. 103(c) and potential 35 U.S.C. 102(e), (f) or (g)
prior art under 35 U.S.C. 103(a).

6.   Claims 1-8, 11-16, 17-24, 27-34, and 38-47 are rejected under 35 U.S.C. 103(a)
as being unpatentable over **Bellegarda et al.** (Article entitled "Exploiting Latent
Semantic Information in Statistical Language Modeling, dated 10/26/2000) and in view
of **Millier et al.** (U.S. Patent 5,899995).

7.   Regarding claim 1, **Bellegarda** teaches a method comprising:

A) mapping the files into a semantic vector space (Page 1279, Abstract);

B) clustering the files within said space (Page 1279, Abstract).

        The examiner notes that **Bellegarda** teaches **"mapping the files into a
semantic vector space"** as "(discrete) words and documents are mapped onto a
(continuous) semantic vector space, in which familiar clustering techniques can be
applied. This leads to the specification of a powerful framework for automatic semantic
classification, as well as the derivation of several language model families with various .
smoothing properties" (Page 1279, Abstract). The examiner further notes that
Bellegarda teaches **"clustering the files within said space"** as "(discrete) words and
documents are mapped onto a (continuous) semantic vector space, in which familiar
clustering techniques can be applied. This leads to the specification of a powerful
framework for automatic semantic classification, as well as the derivation of several
language model families with various smoothing properties" (Page 1279, Abstract).

        Bellegarda does not explicitly teach:

C) displaying the files in a hierarchical format based on the resulting clusters.

        Millier, however, teaches **"displaying the files in a hierarchical format based
on the resulting clusters"** as "The present invention discloses an electronic filing
system for automatically organizing information into a number of storage elements or
folders. The storage elements are folders arranged hierarchically" (Abstract).

        It would have been obvious to one of ordinary skill in the art at the time the
invention was made to combine the teachings of the cited references because teaching
Millier's would have allowed **Bellegarda's** to provide a method for users to find and file

documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 2, **Bellegarda** does not explicitly teach a method comprising:
A) wherein the step of clustering the files is performed as a background routine during . the operation of a computer associated with said file system.

Millier, however, teaches **"wherein the step of clustering the files is performed as a background routine during the operation of a computer associated with said file system"** as "The Reporter 210 may be integrated to specific services...to file stories as they arrive, or to contact the service on a periodic basis to retrieve the relevant documents" (Column 6, lines 25-31).

The examiner notes that it is clear that **Millier's** method has an automated process that locates relevant document in order to place those documents in smart folders.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 3, **Bellegarda** further teaches a method comprising:
A) wherein the step of clustering the files is performed in response to the creation of a new file within the file system (Page 1286, Section: A. Framework Extension).

The examiner notes that **Bellegarda** teaches **"wherein the step of clustering the files is performed in response to the creation of a new file within the file system"** as "finding a new representation for a new document in the space S is straightforward" (Page 1286, Section: A. Framework Extension). The examiner further notes that it is clear that the method of **Bellegarda** clusters when a new document is noticed.

Regarding claim 4, **Bellegarda** further teaches a method comprising:

A) wherein said files are text documents (Page 1279, Abstract); and

B) said mapping is conducted on the basis of a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"wherein said files are text documents"** as "This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches **"said mapping is conducted on the basis of a language model"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Regarding claim 5, **Bellegarda** further teaches a method comprising:

A) wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and

B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said mapping step comprises the steps of constructing a matrix which associates each word in the documents with a vector"** as "The starting point is the construction of a matrix (W) of co-occurrences between words and documents" (Page 1281, Section: A. Feature Extraction) and "The ($M \times N$) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches **"associates each document with a vector"** as "The ($M \times N$)

word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 6, **Bellegarda** further teaches a method comprising:
A) the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"the step of decomposing said matrix to define the words and documents as vectors in a continuous vector space"** as "To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 7, **Bellegarda** further teaches a method comprising:
A) wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches **"wherein said clustering is performed by identifying documents whose vectors are within a threshold distance of one another"** as "This opens up the opportunity to apply familiar clustering techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space" (Page 1286, Section: A. Framework Extension).

Regarding claim 8, **Bellegarda** does not explicitly teach a method comprising:
A) including the step of defining multiple threshold values and clustering said documents in accordance with said multiple threshold values to thereby establish plural levels of clusters.

Millier, however, teaches **"including the step of defining multiple threshold values and clustering said documents in accordance with said multiple threshold values to thereby establish plural levels of clusters"** as "In this example, Input Document 510 satisfies the profiles and constraints of Recent Stories Folder 530 and therefore is stored in Recent Stories Folder 530. Similarly, Input Document 510 satisfies the profiles and constraints of Interesting Stuff-1 Folder 540 and is stored in Interesting-Stuff Folder 540" (Column 8, lines 39-43).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 11, **Bellegarda** teaches a graphical user interface comprising:
A) a virtual file system with a semantic hierarchy, wherein the semantic hierarchy is based on clustering of files based on semantic similarities (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"a virtual file system with a semantic hierarchy, wherein the semantic hierarchy is based on clustering of files based on semantic similarities"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:
B) A graphical user interface configured to display files.

Millier, however, teaches **"A graphical user interface configured to display files"** as "FIG 2A illustrates a screen display 100 on display monitor 62 showing one embodiment of a hierarchical arrangement of the various folders and files in accordance with the teachings of the Smart Folder System of the present invention" (Column 5, lines 39-43).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 13, **Bellegarda** does not explicitly teach a graphical user interface comprising:

A) wherein clustering of the files is initiated by user selection.

Millier, however, teaches **"wherein clustering of the files is initiated by user selection"** as "The Categorize Screen 300 shows an example of a interactive session between the user and the Categorize Dialog process" (Column 7, lines 3-5).

The examiner notes that it is clear that **Millier's** method includes user-initiated commands to begin clustering documents.

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 14, **Bellegarda** further teaches a graphical user interface comprising:

A) wherein clustering of the files is initiated upon creation of a new file in the file system (Page 1286, Section: A. Framework Extension).

The examiner notes that **Bellegarda** teaches **"wherein clustering of the files is initiated upon creation of a new file in the file system"** as "finding a new representation for a new document in the space S is straightforward" (Page 1286, Section: A. Framework Extension). The examiner further notes that it is clear that the method of **Bellegarda** clusters when a new document is noticed.

Regarding claim 15, **Bellegarda** further teaches a graphical user interface comprising:

A) wherein text files are clustered utilizing a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"analyzing files in a file system to determine similarities in data pertaining to their content"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:

B) non-text files are clustered utilizing rule-based techniques.

Millier, however, teaches **"non-text files are clustered utilizing rule-based techniques"** as "The SmartFolder Intelligent Filing System (IFS) allows the user to define a set of rules that allow a document to be filed in multiple contexts of the user's choosing" (Column 3, lines 14-22) and "Graphics Files" (Figure 3).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 16, **Bellegarda** further teaches a graphical user interface comprising:

A) wherein said language model comprises the LSA paradigm (Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches **"wherein said language model comprises the LSA paradigm"** as "The focus of this paper is on semantically driven span extension only, and more specifically on how the LSA paradigm can be exploited to improve statistical language modeling" (Page 1281, Section: D. Organization).

Regarding claim 17, **Bellegarda** teaches a computer-readable media comprising: A) analyzing files in a file system to determine similarities in data pertaining to their content (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"analyzing files in a file system to determine similarities in data pertaining to their content"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:

B) determining a directory structure based on determined similarities between the files;

C) displaying files in hierarchical format based on the determined similarities between the files.

Millier, however, teaches **"determining a directory structure based on determined similarities between the files"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract) and **"displaying files in hierarchical format based on determined similarities between the files"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 18, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said files are text documents (Page 1279, Abstract); and

B) the similarities are based upon the word content of the files (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said files are text documents"** as "This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches **"the similarities are based upon the word content of the files"** as "The starting point is the construction of a matrix (W) of co-occurrences between words and documents" (Page 1281, Section: A. Feature Extraction) and "The ($M \times N$) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 19, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said similarities are determined in accordance with a language model (Page 1279, Abstract, Page 1281, Section: D. Organization); and

B) the files are clustered in accordance with said model (Page 1279, Abstract, Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches **"wherein said similarities are determined in accordance with a language model"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The

examiner further notes that **Bellegarda** teaches **"the files are clustered in accordance with said model"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Regarding claim 20, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said language model comprises the LSA paradigm (Page 1281, Section: D. Organization).

The examiner notes that **Bellegarda** teaches **"wherein said language model comprises the LSA paradigm"** as "The focus of this paper is on semantically driven span extension only, and more specifically on how the LSA paradigm can be exploited to improve statistical language modeling" (Page 1281, Section: D. Organization).

Regarding claim 21, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and

B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said computer-executable code performs the steps of constructing a matrix which associates each word in the documents with a vector"** as "The starting point is the construction of a matrix (W) of co-occurrences between words and documents" (Page 1281, Section: A. Feature Extraction) and "The ($M \times N$) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N,

and each document d_j can be uniquely associated with a column vector of dimension M (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches **"associates each document with a vector"** as "The (M x N) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 22, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said computer-executable code further performs step of decomposing said matrix to define the words and documents as vectors in a continuous vector space"** as "To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 23, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches **"wherein said computer-executable code performs clustering by identifying documents whose vectors are within a threshold distance of one another"** as "This opens up the opportunity to

apply familiar clustering techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space" (Page 1286, Section: A. Framework Extension).

Regarding claim 24, **Bellegarda** does not explicitly teach a computer-readable media comprising:

A) wherein said computer-executable code further performs step of clustering said documents in accordance with multiple threshold values to thereby establish plural levels of clusters.

Millier, however, teaches **"wherein said computer-executable code further performs step of clustering said documents in accordance with multiple threshold values to thereby establish plural levels of clusters"** as "In this example, Input Document 510 satisfies the profiles and constraints of Recent Stories Folder 530 and therefore is stored in Recent Stories Folder 530. Similarly, Input Document 510 satisfies the profiles and constraints of Interesting Stuff-1 Folder 540 and is stored in Interesting-Stuff Folder 540" (Column 8, lines 39-43).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 27, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein the computer executable code performs the following steps: clustering text files within the file system using semantic similarities (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"a semantic hierarchy that is based upon the content of said files"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for

automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:

B) clustering non-text files within the files system using rule-based techniques;

C) labeling the resulting clusters; and

D) displaying the files in a hierarchical format based on the resulting clusters and labels.

Millier, however, teaches **"clustering non-text files within the files system using rule-based techniques"** as "The SmartFolder Intelligent Filing System (IFS) allows the user to define a set of rules that allow a document to be filed in multiple contexts of the user's choosing" (Column 3, lines 14-22) and "Graphics Files" (Figure 3), and **"labeling the resulting clusters"** as "the SmartFolder IFS uses indexing, the indexing is used merely for labeling purposes" (Column 3, lines 35-41), and **"displaying the files in a hierarchical format based on the resulting clusters and labels"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 28, **Bellegarda** teaches a computer system comprising:

A) a file system storing files (Page 1279, Abstract);

C) <u>a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space and cluster the files within said space</u> (Page 1279, Abstract);

D) a semantic hierarchy that is based upon the content of said files (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"a file system storing files"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches **"a processor for analyzing the content of files stored in said file system to map said files into a semantic vector space and cluster the files within said space"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract) The examiner further notes that **Bellegarda** teaches **"a semantic hierarchy that is based upon the content of said files"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:

B) a display device; and

D) a user interface which displays representations of files stored in said file system;

E) wherein said user interface displays said files in accordance with said clustering.

Millier, however, teaches **"a display device"** as "The Categorize Screen 300 shows an example of a interactive session between the user and the Categorize Dialog process" (Column 7, lines 3-5), **"a user interface which displays representations of files stored in said file system"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract), and **"wherein said user interface displays said files in accordance with said**

clustering" as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract)

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 30, **Bellegarda** further teaches a computer system comprising:
A) wherein said files are text documents (Page 1279, Abstract); and
B) said processor maps said files on the basis of a language model (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"wherein said files are text documents"** as "This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches **"said processor maps said files on the basis of a language model"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract).

Regarding claim 31, **Bellegarda** further teaches a computer system comprising:
A) wherein said processor constructs a matrix which associates each word in the documents with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition); and
B) associates each document with a vector (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said processor constructs a matrix which associates each word in the documents with a vector"** as "The starting point is the construction of a matrix (W) of co-occurrences between words and documents" (Page 1281, Section: A. Feature Extraction) and "The ($M \times N$) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M (Page 1281, Section: B. Singular Value Decomposition). The examiner further notes that **Bellegarda** teaches **"associates each document with a vector"** as "The ($M \times N$) word-document matrix W resulting from the above feature extraction defines two vector representations for the words and the documents. Each word ω_i can be uniquely associated with a row vector of dimension N, and each document d_j can be uniquely associated with a column vector of dimension M" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 32, **Bellegarda** further teaches a computer-readable media comprising:

A) wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space (Page 1281, Section: A. Feature Extraction, Section: B. Singular Value Decomposition).

The examiner notes that **Bellegarda** teaches **"wherein said processor further decomposes said matrix to define the words and documents as vectors in a continuous vector space"** as "To address these issues, it is useful to employ a singular value decomposition (SVD), a technique closely related to eigenvector decomposition and factor analysis" (Page 1281, Section: B. Singular Value Decomposition).

Regarding claim 33, **Bellegarda** further teaches a computer system comprising:

A) wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another (Page 1284, Section: A. Word Clustering).

The examiner notes that **Bellegarda** teaches **"wherein said processor clusters the files by identifying documents whose vectors are within a threshold distance of one another"** as "This opens up the opportunity to apply familiar clustering techniques in S, as long as a distance measure consistent with the SVD formalism is defined on the vector space" (Page 1286, Section: A. Framework Extension).

Regarding claim 34, **Bellegarda** does not explicitly teach a computer system comprising:

A) wherein said processor clusters said files in accordance with multiple threshold values to thereby establish plural levels of clusters.

Millier, however, teaches **"wherein said processor clusters said files in accordance with multiple threshold values to thereby establish plural levels of clusters"** as "In this example, Input Document 510 satisfies the profiles and constraints of Recent Stories Folder 530 and therefore is stored in Recent Stories Folder 530. Similarly, Input Document 510 satisfies the profiles and constraints of Interesting Stuff-1 Folder 540 and is stored in Interesting-Stuff Folder 540" (Column 8, lines 39-43).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 37, **Bellegarda** does not explicitly teach a method comprising:
A) wherein clustering includes organizing the clusters into a hierarchical directory structure.

Millier, however, teaches **"wherein clustering includes organizing the clusters into a hierarchical directory structure"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of

storage elements or folders. The storage elements are folders arranged hierarchically"
(Abstract).

It would have been obvious to one of ordinary skill in the art at the time the
invention was made to combine the teachings of the cited references because teaching
Millier's would have allowed **Bellegarda's** to provide a method for users to find and file
documents in multiple locations in order to gain a much more natural mode of
information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 38, **Bellegarda** teaches a method comprising:
A) mapping all words of the plurality of documents and the plurality of documents in a
semantic vector space (Page 1279, Abstract);
B) clustering the plurality of documents to a plurality of clusters based on semantic
similarities of the plurality of documents (Page 1279, Abstract).

The examiner notes that **Bellegarda** teaches **"mapping all words of the
plurality of documents and the plurality of documents in a semantic vector
space"** as "(discrete) words and documents are mapped onto a (continuous) semantic
vector space, in which familiar clustering techniques can be applied. This leads to the
specification of a powerful framework for automatic semantic classification, as well as
the derivation of several language model families with various smoothing properties"
(Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches
**"clustering the plurality of documents to a plurality of clusters based on semantic
similarities of the plurality of documents"** as "(discrete) words and documents are
mapped onto a (continuous) semantic vector space, in which familiar clustering
techniques can be applied. This leads to the specification of a powerful framework for
automatic semantic classification, as well as the derivation of several language model
families with various smoothing properties" (Page 1279, Abstract).

Bellegarda does not explicitly teach:
C) outputting the plurality of documents in a hierarchical format based on a result of
clustering the plurality of documents.

Millier, however, teaches **"outputting the plurality of documents in a hierarchical format based on a result of clustering the plurality of documents"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 39, **Bellegarda** further teaches a method comprising:
A) wherein the step of clustering the plurality of documents comprises: generating the plurality of clusters based on the semantic similarities of the plurality of documents (Page 1279, Abstract);
B) organizing the plurality of clusters into one or more granularity levels (Page 1294);
D) determining a membership of each of the plurality of documents to a particular cluster based on a content of each document in relationship to the particular cluster (Page 1279, Abstract, Page 1286, Figure 5).

The examiner notes that **Bellegarda** teaches **"wherein the step of clustering the plurality of documents comprises: generating the plurality of clusters based on the semantic similarities of the plurality of documents"** as "(discrete) words and documents are mapped onto a (continuous) semantic vector space, in which familiar clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The examiner further notes that **Bellegarda** teaches **"organizing the plurality of clusters into one or more granularity levels"** as "In addition, well-known clustering algorithms can be applied efficiently, which allows the characterization of parallel layers of semantic knowledge in the space with variable granularity" (Page 1294). The examiner

further notes that **Bellegarda** teaches **"determining a membership of each of the plurality of documents to a particular cluster based on a content of each document in relationship to the particular cluster"** as "This paper focuses on the use of latent semantic analysis, a paradigm that automatically uncovers the salient semantic relationships between words and documents in a given corpus" (Page 1279, Abstract), "The vector v_P, indeed seen to be functionally similar to a document vector, corresponds to the representation of the new document in the space . This is illustrated in Fig. 5, which depicts the R=2 rendition of S, with x's and o's denoting the original words and documents, respectively, and the symbol "@" showing the location of the new document" (Page 1286) and "Once the representation (15) is obtained, the "closeness" between the new document and any document cluster Fig. 1. The row vectors of (i.e., words) are projected onto the can then be expressed as calculated from (13) in the previous section" (Page 1286).

　　　　Bellegarda does not explicitly teach:

C) wherein the plurality of clusters are organized as directories in the hierarchical format.

　　　　Millier, however, teaches **"wherein the plurality of clusters are organized as directories in the hierarchical format"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract).

　　　　It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

　　　　Regarding claim 40, **Bellegarda** further teaches a method comprising:

A) wherein the step of generating the plurality of clusters comprises: generating a matrix W of dimensions (MxN), wherein M is a number of a list of words and symbols

that occur in the plurality of documents, N is a number of the plurality of documents, each entry $w_{i,j}$ of the matrix W is pf the form $w_{i,j} = (1-\varepsilon_i)(c_{i,j}/n_j)$, $c_{i,j}$ is a number of times a word w_i occurs in a document d_j, n_j is a total number of words in the document d_j, and c_i is a normalized entropy of the word w_i in a corpus T of the collection of the N documents (Page 1281, Section: Feature Extraction);

B) performing a singular value decomposition on the matrix W such that $W=USV^T$, wherein U is a (MxR) left singular matrix with row vectors $u_i(1<=i<=M)$ representing word vectors, S is a diagonal matrix of singular values $s_1>= s_2>=...>= s_R>0$, V is a (NxR) right singular matrix with row vectors $v_i(1<=i<=N)$ representing document vectors, T denotes matrix transposition, and R<<M,N (Page 1281, Section: Singular Value Decomposition);

C) evaluating closeness of the plurality of documents based on the document vectors (Page 1284: Section: Word Clustering); and

D) merging semantic information of the plurality of documents based on the closeness of the documents (Page 1284: Section: Word Clustering).

The examiner notes that **Bellegarda** clearly teaches the formulas on page 1281.

Regarding claim 41, **Bellegarda** further teaches a method comprising:

A) wherein the step of evaluating the closeness of the plurality of the documents comprises: measuring closeness of two documents v_j, v_k according to an equation $K(v_j, v_k)=cos((v_jS, v_kS)= (v_jS^2 v_kT)/((v_jS)(v_kS))$ (Page 1284: Section: Word Clustering).

The examiner notes that **Bellegarda** clearly teaches the formula on page 1284.

Regarding claim 42, **Bellegarda** further teaches a method comprising:

A) controlling a number of clusters based on a predetermined threshold of a variance of a cluster resulting from performing the step of merging the semantic information (Pages 1283-1284);

B) wherein the variance σ^2 of the cluster is measured according to an equation $\sigma^2=[(n_1\sigma_1^2+ n_2\sigma_2^2)/(n_1+ n_2)]+[(n_1 n_2)((u_1+u_2)^2)/(n_1+ n_2)]$, u_1, σ_1^2, and u_2, σ_2^2, are means

and variances of candidates for merging and n_1 and n_2 are sizes of the clusters (Pages 1283-1284)

Regarding claim 43, **Bellegarda** further teaches a method comprising:

A) wherein the step of generating the plurality of clusters further comprises: clustering the plurality of documents into the plurality of clusters based on the predetermined threshold variance (Page 1284);

B) grouping the plurality of clusters to a one or more super clusters based on one or more additional predetermined threshold variances, wherein super cluster includes one or more lower level clusters and/or super clusters (Page 1284).

Bellegarda does not explicitly teach:

C) wherein each cluster and super cluster is represented as a directory and each super cluster directory is hierarchically above each lower level cluster directory and/or super cluster directory.

Millier, however, teaches **"wherein each cluster and super cluster is represented as a directory and each super cluster directory is hierarchically above each lower level cluster directory and/or super cluster directory"** as "The present invention discloses an electronic filing system for automatically organizing information into a number of storage elements or folders. The storage elements are folders arranged hierarchically" (Abstract).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Millier's** would have allowed **Bellegarda's** to provide a method for users to find and file documents in multiple locations in order to gain a much more natural mode of information access, as noted by **Millier** (Column 1, lines 65-67).

Regarding claim 44, **Bellegarda** further teaches a method comprising:

A) determining whether one or more plurality of documents fall outside of the threshold variance of the lowest level of clusters (Page 1284, Abstract);

B) determining whether the one or more of the plurality of documents that fall outside the thresholds of the lowest level of clusters fall within the threshold variance of a particular super cluster (Page 1284); and

C) grouping the one or more of the plurality of documents to be included in the particular super cluster when it is determined that the one or more of the plurality of documents fall within the threshold variance of the particular super cluster (Page 1284).

Regarding claim 45, **Bellegarda** further teaches a method comprising:

A) for each cluster, determining a closeness of the words in the documents of the cluster to the cluster (Page 1288);

B) choosing one or more words based on the closeness of the one or more words to the cluster and assigning the one ore more words as a label to the cluster (Page 1285).

The examiner further notes that **Bellegarda** teaches **"choosing one or more words based on the closeness of the one or more words to the cluster and assigning the one ore more words as a label to the cluster"** as "While clearly not matching the hand-labeling, LSA document clustering in this example still seems reasonable. In particular, as one would expect, he distribution for the natural science topic is relatively close to the distribution for the applied science topic (see the two solid lines), but quite different from the two other topic distributions (in dashed lines). From that standpoint, the data-driven LSA clusters appear to adequately cover the semantic space" (Page 1285).

Regarding claim 46, **Bellegarda** further teaches a method comprising:

A) wherein the step of determining the closeness of words and documents in the cluster comprises: measuring the closeness of a word vector u_i and a document vector v_k, according to an equation $K(u_i, v_k)=cos((u_iS^{1/2}, v_kS^{1/2})= (u_iSv_kT)/((u_iS^{1/2})(v_kS^{1/2}))$ (Page 1288).

The examiner notes that **Bellegarda** clearly teaches the formula on page 1288.

Regarding claim 47, **Bellegarda** further teaches a method comprising:

A) wherein the step of choosing the one or more words comprises choosing all words that are within a predetermined closeness threshold K (Pages 1284, 1288).

The examiner notes that **Bellegarda** teaches **"wherein the step of choosing the one or more words comprises choosing all words that are within a predetermined closeness threshold K"** as "Each supercluster is then itself partitioned using bottom-up clustering, resulting in a final set of clusters C_k 1<=k<=K" (Page 1284)

8.　　　Claims 9-10, 25-26, and 35-36 are rejected under 35 U.S.C. 103(a) as being unpatentable over **Bellegarda et al.** (Article entitled "Exploiting Latent Semantic Information in Statistical Language Modeling, dated 10/26/2000) and in view of **Millier et al.** (U.S. Patent 5,899995) as applied to claims 1-8, 11-16, 17-24, 27-34, and 38-47 and further in view of **Kusama** (U.S. Patent 7,085,767).

9.　　　Regarding claim 9, **Bellegarda** and **Millier** do not explicitly teach a method comprising:

A) including the step of automatically labeling the clusters <u>based on the resulting clusters</u>.

Kusama, however teaches **"including the step of automatically labeling the clusters <u>based on the resulting clusters</u>"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by automatically devising a folder name in order to lessen the burden of having to conform to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Regarding claim 10, **Bellegarda** and **Millier** do not explicitly teach a method comprising:

A) wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

Kusama, however teaches **"wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by automatically devising a folder name in order to lessen the burden of having to conform to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Regarding claim 25, **Bellegarda** and **Millier** do not explicitly teach a computer-readable media comprising:

A) wherein said computer-executable code performs step of automatically labeling the clusters <u>based on the resulting clusters</u>.

Kusama, however teaches **"wherein said computer-executable code performs step of automatically labeling the clusters <u>based on the resulting clusters</u>"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by automatically devising a folder name in order to lessen the burden of having to conform to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Regarding claim 26, **Bellegarda** and **Millier** do not explicitly teach a computer-readable media comprising:

A) wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

Kusama, however teaches **"wherein said labeling comprises selecting representative words based on the closeness of their vectors to the document vectors in a cluster"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by automatically devising a folder name in order to lessen the burden of having to conform to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Regarding claim 35, **Bellegarda** and **Millier** do not explicitly teach a computer system comprising:

A) wherein said processor automatically labels the clusters based on the resulting clusters.

Kusama, however teaches **"wherein said processor automatically labels the clusters <u>based on the resulting clusters</u>"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by automatically devising a folder name in order to lessen the burden of having to conform to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Regarding claim 36, **Bellegarda** and **Millier** do not explicitly teach a computer system comprising:

A) wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster.

Kusama, however teaches **"wherein said processor labels the clusters by selecting representative words based on the closeness of their vectors to the document vectors in a cluster"** as "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53).

It would have been obvious to one of ordinary skill in the art at the time the invention was made to combine the teachings of the cited references because teaching **Kusama's** would have allowed **Bellegarda's** and **Millier's** to provide a method for users to find and file documents in multiple locations in order to generate/copy data by

automatically devising a folder name in order to lessen the burden of having to conform
to the content of the data, as noted by **Kusama** (Column 1, lines 34-38).

Response to Arguments

10.   Applicant's arguments filed on 03/21/2007 have been fully considered but they
are not persuasive.

Applicant argues on page 15, that "**In other words, the creation of the
SmartFolders themselves is a completely manual process performed by the user.
In Millier, the clustering of the files is determined based on the hierarchical
structure that is pre-existing. In contrast, as recited in claim 1, the hierarchal
format is based on the clustering of the files. Thus, Millier directly teaches away
from the feature as recited**". However, the examiner wishes to state that **Bellegarda**
teaches the aforementioned limitation (as stated on page 10 of the non-final office
action mailed on 12/21/2006). Moreover, the examiner wishes to point to page 1279 of
Bellegarda which states "(discrete) words and documents are mapped onto a
(continuous) semantic vector space, in which familiar clustering techniques can be
applied. This leads to the specification of a powerful framework for automatic semantic
classification, as well as the derivation of several language model families with various
smoothing properties" (Page 1279, Abstract). The examiner further wishes to state that
Bellegarda's method clearly is the same method as taught by the instant application.
Moreover, the examiner further wishes to state that clustering techniques of **Bellegarda**
clearly create the space to enable documents and/or words to be grouped together
based on semantic similarities.

Applicant argues on page 15, that "**Independent claim 11 recites, in part
"wherein the semantic hierarchy is based on clustering of files based on
semantic similarities." In other words, the hierarchy is based on the clustering of
the files, whereas Millier teaches exactly the opposite**". However, the examiner
wishes to state that **Bellegarda** teaches the aforementioned limitation (as stated on
page 10 of the non-final office action mailed on 12/21/2006). Moreover, the examiner
wishes to point to page 1279 of **Bellegarda** which states "(discrete) words and
documents are mapped onto a (continuous) semantic vector space, in which familiar

clustering techniques can be applied. This leads to the specification of a powerful framework for automatic semantic classification, as well as the derivation of several language model families with various smoothing properties" (Page 1279, Abstract). The examiner further wishes to state that **Bellegarda's** method clearly is the same method as taught by the instant application. Moreover, the examiner further wishes to state that clustering techniques of **Bellegarda** clearly create the space to enable documents and/or words to be grouped together based on semantic similarities.

Applicant argues on page 16, that **"claims 9, 25, and 25 all recite that the step of automatically labeling the clusters, based on the resulting clusters. None of Bellegarda, Millier, and Kusama teach or suggest this feature"**. However, the examiner wishes to point to Column 5 of **Kusama** which states "the "Title" of "cardinfo.xml" is read, and the folder having the same name as the meta data being saved in the "Title" are generated at a predetermined location in the binary data storage device. According to this processing, in the case where this, for example, the meta data "cardinfo.xml" depicted in FIG. 10, then the folder having the name of "Party" which is written in the "Title" is generated" (Column 5, lines 46-53). The examiner further wishes to state that it is clear that **Kusama** automatically labels and names clusters.

Conclusion

11.   The prior art made of record and not relied upon is considered pertinent to applicant's disclosure.

U.S. Patent 6,820,094 issued to **Ferguson et al.** on 16 November 2004. The subject matter disclosed therein is pertinent to that of claims 1-36 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. PGPUB 2004/0249865 issued to **Lee et al.** on 09 December 2004. The subject matter disclosed therein is pertinent to that of claims 1-36 (e.g., methods to automatically name and label folders).

U.S. PGPUB 2004/0148453 issued to **Watanabe et al.** on 29 July 2004. The subject matter disclosed therein is pertinent to that of claims 1-36 (e.g., methods to automatically name and label folders).

U.S. Patent 5,819,258 issued to **Vaithyanathan et al.** on 06 October 1998. The subject matter disclosed therein is pertinent to that of claims 1-36 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

U.S. Patent 6,360,227issued to **Aggarwal et al.** on 19 March 2002. The subject matter disclosed therein is pertinent to that of claims 1-36 (e.g., methods to use to smart folders to automatically organize and relate relevant files).

11.   Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Contact Information

12.   Any inquiry concerning this communication or earlier communications from the examiner should be directed to Mahesh Dwivedi whose telephone number is (571) 272-2731. The examiner can normally be reached on Monday to Friday 8:20 am – 4:40 pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Tim Vo can be reached (571) 272-3642. The fax number for the organization where this application or proceeding is assigned is (571) 273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should
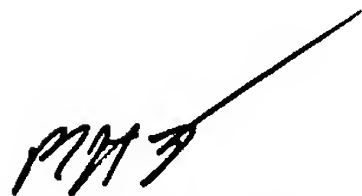
you have questions on access to the Private PAIR system, contact the Electronic
Business Center (EBC) at 866-217-9197 (toll-free).

Mahesh Dwivedi

Patent Examiner

Art Unit 2168

June 07, 2007

TIM VO
SUPERVISORY PATENT EXAMINER
TECHNOLOGY CENTER 2100